**TriAD**: Capturing harmonics with 3D Convolutions

Deep learning automatic music transcription systems outperformed previous ones fully based on manual feature design, at the cost of being computationally expensive. New trends move towards smaller models that maintain such results by embedding musical knowledge in the network architecture. We present TriAD, a convolutional block that achieves an unequally distanced dilation over the frequency axis, allowing our method to capture multiple harmonics with a single yet small kernel, in contrast to existing methods.

*When a note is played, the fundamental frequency sounds along its harmonics: 2nd, 3rd, 4th, etc.*

*Automatic transcription systems uses that pattern to obtain accurate pitch information.*

*The challenge: harmonics are not equally distanced over the frequency axis.*

---

**TriAD**

- **Capturing harmonics**
- **with 3D Convolutions**

**HPPNet** [1] as the reference model.

**Tested multiple different harmonic blocks**

**Training on MAESTRO** [2]

**Evaluation on MAESTRO and MAPS** [3]

**SOTA with less parameters.**

**Better correlation with harmonic information (See the table below)**

### Table 1: Performance Comparison

<table>
<thead>
<tr>
<th>Block</th>
<th>MAESTRO</th>
<th>MAPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TriAD (Ours)</td>
<td>90.14%</td>
<td>71.58%</td>
</tr>
<tr>
<td>HD-Conv [1]</td>
<td>84.89%</td>
<td>69.96%</td>
</tr>
</tbody>
</table>

---

- The input to the network must be a log-frequency representation of the spectrum.
- The CQT is split into an octave/pitch spectrogram.
- 3D are kernels dilated at the pitch-class dimension. Certain pitch-class intervals are associated with certain harmonics.
  - E.g. 2nd and 4th harmonics are octaves; 3rd and 6th harmonics are perfect fifths.
- The output of the kernels convolutions get aggregated.

---

